RANGEL, DENISE A., M.A. Elliptic Curves and Factoring. (2010)
Directed by Dr. Paul Duvall. 47 pp.

The Elliptic Curve Method (ECM) is a powerful and widely used algorithm for factorization which can be implemented with several different forms of elliptic curves. We will give some general background on the theory of elliptic curves and the ideas behind ECM. We will then discuss three families of curves and compare the speed of their addition methods in the implementation of ECM.

ELLIPTIC CURVES AND FACTORING

by

Denise A. Rangel

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
2010

Approved by

Committee Chair

APPROVAL PAGE

This thesis has been approved by the following committee of the
Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair_____

Committee Members_____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

ACKNOWLEDGMENTS

I would like to give my deepest gratitude to Dr. Paul Duvall for his guidance and advice throughout this process. I would also like to thank Dr. Maya Chhetri and Dr. Dan Yasaki for their support and helpful suggestions.

TABLE OF CONTENTS

Page

CHAPTER

I. INTRODUCTION 1

II. BASIC INFORMATION ON ELLIPTIC CURVES 3

2.1. Projective Space 6
2.2. Elliptic Curve Addition 8
2.3. The j-invariant 12

III. ELLIPTIC CURVES OVER FINITE FIELDS 16

IV. THE ELLIPTIC CURVE FACTORING METHOD 18

4.1. Phase 1 19
4.2. Phase 2 20

V. MONTGOMERY CURVES 23

5.1. Montgomery Addition 26
5.2. Implementation of ECM 29

VI. EDWARDS CURVES 34

6.1. Edwards Addition 35

VII. CONCLUSION 38

BIBLIOGRAPHY 39

APPENDIX A. CODE 41

CHAPTER I

INTRODUCTION

The Elliptic Curve Factoring Method (ECM) was first developed more than twenty years ago. Since that time many people have studied it, and thus improved upon it. The choice of which form of elliptic curves to use in the ECM has often been a topic of discussion. There is the standard form called Weierstrass equation which has a short and a long form. We will use it to define a geometrically motivated addition method for the points on an elliptic curve, which will lead us into using elliptic curves to define a group.

The next type of elliptic curves we will look at are Montgomery curves, which first appeared soon after the ECM did. In addition to introducing a new form of curves, a second phase of the ECM was introduced which improves the chance of algorithm succeeding. A primary purpose for Montgomery curves was use in the ECM since its addition method is only defined for adding a multiple of a point to a multiple of the same point. However, unlike the Weierstrass form, it does not require inversion in its addition method. Thus the use of Montgomery curves in the ECM leads to a faster algorithm.

The final type of elliptic curves we will explore are called Edwards curves and are a recent development. The Edwards form does have addition defined for two different points on a curve. However the addition formula is quite long but certain parameters can be defined for it so we can increase the computing speed of it. To decide which form, the Montgomery or Edwards, would be better to use in the ECM we will compare the speed of the addition methods in computing a large

multiple of a point. We will also briefly discuss a method for choosing a bound in the ECM if we are looking for a certain size factor. However, before we discuss the Montgomery and Edwards curves, and the ECM we need to have some general background on elliptic curves.

CHAPTER II

BASIC INFORMATION ON ELLIPTIC CURVES

Let K be any field where the characteristic is not 2 or 3, and let $\overline{K}$ be the algebraic closure of K. In algebraic geometry an elliptic curve is defined as a quite complex object, but for our proposes we define it as follows

Definition 1. *(Elliptic Curve) An elliptic curve, E is a set of points $(x, y) \in K \times K$ satisfying the equation*

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \tag{2.1}$$

*where $\{a_1 \dots a_6\} \in K$, along with a point at infinity, $\mathcal{O}$. We say that E is **defined over** K, and write $E(K)$.*

For now just consider the point of infinity as extra point on the curve. It will be defined in greater detail in section 2.1. The equation (2.1) is called the *General Weierstrass Equation*, and is most useful when working over fields where the characteristic is 2 or 3. However if the field is not of characteristic 2 or 3 then it can be changed into the *Weierstrass form* (often called the short form) which is defined as

$$y^2 = x^3 + Ax + B \tag{2.2}$$

where $A, B \in K$.

For technical reason we want the cubic, $x^3 + Ax + B$ to have distinct roots over $\overline{K}$, the algebraic closure of K. Therefore we assume that the discriminant of the cubic is not equal to zero. For elliptic curves in the Weierstrass form this gives us the condition that $4A^3 + 27B^2 \neq 0$.

Since the Weierstrass form is derived from the general Weierstrass equation, it is possible to change curves from the general form to the short form. Given any in the long Weierstrass form to the short form the following change of variable can be used, and since the characteristic of K is not 2, it will always possible to divide by 2 during the process of completing the square. So for

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

rearrange to

$$\left(y + \frac{a_1x}{2} + \frac{a_3}{2}\right)^2 = x^3 + \left(a_2 + \frac{a_1^2}{4}\right)x^2 + \left(a_4 + \frac{a_1a_3}{2}\right)x + \left(\frac{a_3^2}{4} + a_6\right).$$

This can be rewritten as

$$y_1^2 = x^3m + a_2'x^2 + a_4'x + a_6'.$$

For $y_1 = y + \frac{a_1}{2}x + \frac{a_3}{2}$ and where a_2', $a_4'$and a_6' are the constant coefficients to their respective variables. Also since the characteristic is not 3, we can let $x_1 = x + \frac{a_2'}{3}$, the Weierstrass form of $y_1^2 = x_1^3 + Ax_1 + B$ can now be obtained.

Another thing to notice is that the coefficients for y^2 and x^3 are always assumed to be one in the Weierstrass form. If the coefficients are desired to be one and they are not, another change of variables can used to obtain a curve in the Weierstrass form. Suppose that $c, d \neq 0$ and that $cy^2 = dx^3 + ax + b$. Now multiply both sides by c^3d^2 to get

$$c^4d^2y^2 = c^3d^3x^3 + ac^3d^2 + bc^3d^2$$

$$(c^2dy)^2 = (cdx)^3 + (ac^2d)(cdx) + (bc^3d^2).$$

Define $y_1 = c^2dy$ and $x_1 = cdx$ to obtain the Weierstrass form of the equation

$$y_1^2 = x_1^3 + Ax_1 + B$$

where $A = ac^2d$ and $B = bc^3d^2$.

Suppose we know two points on an elliptic curve and want to use them to find a third. This is possible if we use the line formed by connecting the two points and then find where that line intersects the curve in a third point. Take for example the curve E defined by $y^2 = x^3 + 5x + 19$ which has points $P = (1, 5)$ and $Q = (-2, 1)$ on it. The following fact will be helpful in finding the third point.

Proposition 1. *For a monic cubic equation over K, the coefficient of the x^2 term is equal to the negative of the sum of its roots.*

Proof. Let $f(x) = x^3 + \alpha_2 x^2 + \alpha_1 x + \alpha_0$ for $\alpha_2, \alpha_1, \alpha_0 \in K$ and let r_1, r_2, r_3 be the roots of $f(x)$. Then

$$\begin{aligned} f(x) &= (x - r_1)(x - r_2)(x - r_3) \\ &= x^3 - (r_1 + r_2 + r_3)x^2 + (r_1r_2 + r_1r_3 + r_2r_3)x - r_1r_2r_3 \\ &= x^3 + \alpha_2 x^2 + \alpha_1 x + \alpha_0 \end{aligned}$$

Therefore, $\alpha_2 = -(r_1 + r_2 + r_3)$

□

Now to find another point on E, we first need to find the line (call it PQ) that contains both P and Q, and then find the intersection of it with E. The slope of PQ is $m = \frac{4}{3}$, and therefore PQ is the line $y = \frac{4}{3}x + \frac{11}{3}$. Substituting PQ for y in E we get

$$\left(\frac{4}{3}x + \frac{11}{3}\right)^2 = x^3 + 5x + 19$$

$$\frac{16}{9}x^2 + \frac{44}{9}x + \frac{121}{9} = x^3 + 5x + 19$$

$$0 = x^3 - \frac{16}{9}x^2 - \frac{43}{9}x - \frac{40}{9}.$$

Now the goal is to factor this cubic, which is made extremely easier since we already know two of the roots, namely the x-coordinates of P and Q. Therefore

$$x^3 - \frac{16}{9}x^2 - \frac{43}{9}x - \frac{40}{9} = (x-1)(x+2)(x-r).$$

and from Proposition 1

$$\frac{16}{9} = 1 - 2 + r$$

$$r = \frac{25}{9}$$

Plugging r into the equation for the line, a new point $R = (\frac{25}{9}, \frac{199}{27})$ is obtained. We have actually found two points on the curve. For if some point (x_1, y_1) satisfies an elliptic curve, $y^2 = x^3 + Ax + B$ then so will the point $(x_1, -y_1)$. Therefore for every point (x, y) on an elliptic curve, the point $(x, -y)$ will also be on the curve. In the above example R was found as well as $S = (\frac{25}{9}, -\frac{199}{27})$. This process can be repeated with different points, to create different lines, and find more points on the curve.

In general we want to define a way to add two points on an elliptic curve to get a third one. In the above example, we will define addition to be $P + Q = S$. Recall that S was found by taking the third point R on the line formed by P and Q and then reflecting R across the x-axis. However to define addition explicitly, we first need to bring in the idea of a point at infinity and projective space.

2.1 Projective Space

Projective space can be defined for any finite dimension, but for our purposes we only need to define it for the two dimensional case.

Definition 2. *(Projective plane) The two-dimensional* projective plane, $\mathbb{P}^2_{\mathrm{K}}$ *over a field K is the set of equivalent classes of triples (x, y, z) where x, y, $z \in K$, not all equal to zero. Two triples (x, y, z) and (a, b, c) are* equivalent *if there exists a $\gamma \in K$, not equal to zero, such that $(x, y, z) = (\gamma a, \gamma b, \gamma c)$ and we write $(x, y, z) \sim (a, b, c)$.*

Since this type of an equivalence class is dependent on the ratios of x, y, and z, it is written as $(x : y : z)$. If we have a triple $(x : y : z)$ and $z \neq 0$ then we can divide through by z

$$(x : y : z) = (\frac{x}{z} : \frac{y}{z} : 1).$$

These types of triples in $\mathbb{P}^2_{\mathrm{K}}$ where $z \neq 0$ can be identified as the finite points in the space K^2. For example, if the field is $\mathbb{R}$, then the triple $(7 : 3 : 1) \in \mathbb{P}^2_{\mathbb{R}}$ can be thought of as corresponding to the point $(7, 3)$ in the real plane. The collection of these finite points are called the affine plane.

Definition 3. *(Affine Plane) $\mathbb{A}^2_K$, the affine plane, is the set of pairs, $(x, y) \in K \times K$. We think of $\mathbb{A}^2_K$ as a subset of $\mathbb{P}^2_{\mathrm{K}}$ where $(x, y) \to (x : y : 1)$.*

The rest of the points in Projective space are called the points of infinity.

Definition 4. *(Points at infinity) The points in $\mathbb{P}^2_{\mathrm{K}}$ where $z = 0$ are called the* points of infinity.

To describe these points of infinity on a curve, the equation must first be homogeneous.

Definition 5. *(Homogeneous) A polynomial is* homogeneous of degree d *if for every term of the polynomial, the sum of the exponents of the variables is equal to d.*

For example the polynomial $f(x, y, z,) = 5x^4 + 3x^2yz + xyz^2$ is homogeneous of degree 4. A polynomial (in the plane is given in terms of x and y) can be put

into a homogeneous form by simply inserting the powers of z that are needed. For example, if $g(x, y) = x^3 + 3x^2 - 8xy + 2$ then it is made homogeneous by inserting z's so that the sum of the degrees of each term is 3,

$$g(x, y, z) = x^3 + 3x^2z - 8xyz + 2z^3.$$

Then $g(x, y, 1) = g(x, y)$. So for any point (x, y) in the affine plane, it corresponds to the point $(x, y, 1)$ in the projective plane. The homogeneous form for elliptic curves in the Weierstrass form is

$$y^2z = x^3 + Axz^2 + Bz^3. \tag{2.3}$$

To find the point of infinity for equation 2.3 set $z = 0$, then

$$\begin{aligned} y^2z &= x^3 + Axz^2 + Bz^3 \\ y^2(0) &= x^3 + Ax(0)^2 + B(0)^3 \\ 0 &= x^3 \quad \Rightarrow \quad x = 0 \end{aligned}$$

So we have the point $(0, y, 0) \sim (0, 1, 0)$. This will be the point at infinity for elliptic curves, when written with projective coordinates. If strictly in the affine plane then the point of infinity is denoted as $\mathcal{O}$. For most of our purposes we will be working in the Affine plane. However the use of projective coordinate will become very crucial when we discuss Montgomery curves.

2.2 Elliptic Curve Addition

Let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ be points on an elliptic curve, E over the field K, defined as $y^2 = x^3 + Ax + B$. Define $P_1 + P_2 = P_3$ as follows:

Case 1. $P_1 \neq P_2$, $x_1 \neq x_2$, and neither are $\mathcal{O}$

Make the line, L that contains both P_1 and P_2. This gives us the slope

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

which is well defined since $x_1 \neq x_2$. Then L has the form

$$y = m(x - x_1) + y_1$$

Now substitute into E to find the intersection of L and E.

$$y^2 = x^3 + Ax + B$$

$$(m(x - x_1) + y_1)^2 = x^3 + Ax + B$$

$$m^2(x - x_1)^2 - 2m(x - x_1)y_1 + y_1^2 = x^3 + Ax + B$$

$$m^2(x^2 - 2xx_1 + x_1^2) - 2m(x - x_1)y_1 + y_1^2 = x^3 + Ax + B$$

$$m^2x^2 - 2m^2x_1 + m^2x_1^2 - 2mxy_1 + 2mx_1y_1 + y_1^2 = x^3 + Ax + B$$

$$0 = x^3 - m^2x^2 + (A + 2m^2x_1 + 2my_1)x + (B - m^2x_1 - 2mx_1y_1 - y_1^2)$$

The three roots of this cubic will gives us the x values of the points of intersection, two of which we already know. Since in monic cubic polynomials, the opposite of the sum of the roots is the coefficient of x^2, (Proposition 1.)

$$m^2 = x_1 + x_2 + x'$$

$$x' = m^2 - x_1 - x_2$$

So then,

$$y' = m(x' - x_1) + y_1$$

This gives us the third point, (x', y'), on $E \cap L$. However, the point we want to define as the sum of P_1 and P_2 is its reflection across the x-axis, $(x', -y')$. Therefore, for $P_3 = (x_3, y_3)$

$$P_1 + P_2 = P_3$$

where

$$x_3 = m^2 - x_1 - x_2 \qquad y_3 = m(x_1 - x_3) - y_1$$

Case 2. $P_1 \neq P_2$ but $x_1 = x_2$

Then the line containing P_1 and P_2 is vertical, so it intersects E at $\mathcal{O}$. If $\mathcal{O}$ is reflected across the x-axis it is still $\mathcal{O}$. Therefore

$$P_1 + P_2 = \mathcal{O}$$

Case 3. $P_1 = P_2 = (x_1, y_1)$ (Doubling)

Then the line we want is the line tangent to the curve at P_1. (Here we assume that $y_1 \neq 0$, because if $y_1 = 0$ then the tangent line is vertical and follows case 2.) Implicit differentiation allows us to find the slope of this tangent line L.

$$2y\frac{dy}{dx} = 3x^2 + A$$

$$\mu = \frac{dy}{dx} = \frac{3x_1^2 + A}{2y_1}$$

so the line is defined as

$$y = \mu(x - x_1) + y_1.$$

Proceeding as before, we obtain the following monic cubic equation

$$0 = x^3 - \mu^2 x^2 + (A + 2\mu^2 x_1 + 2\mu y_1)x + (B - \mu^2 x_1 - 2\mu x_1 y_1 - y_1^2).$$

Now we only know one of the roots, x_1, but this is a double root since a tangent line was used, so we can proceed just as before obtaining, $P_3 = (x_3, y_3)$ where

$$x_3 = \mu^2 - 2x_1 \quad \text{and} \quad y_3 = \mu(x_1 - x_3) - y_1$$

Case 4. $P_2 = \mathcal{O}$

Then the line through P_1 and P_2 is vertical. This line intersects E at P_1's reflection across the x-axis, the point $(x_1, -y_1)$ and reflecting it again simply gives us P_1. Therefore,

$$P_1 + \mathcal{O} = P_1$$

Theorem 1. *For an elliptic curve E, the addition of points on E as defined above satisfies the following:*

1. *For all P_1, P_2 on E: $P_1 + P_2 = P_2 + P_1$ (Commutative).*
2. *For all points, P on E: $P + \mathcal{O} = P$ (Identity element).*
3. *For any P on E there exist a $(-P)$ on E such that $P + (-P) = \mathcal{O}$ (Inverses)*
4. *For all P_1, P_2, P_3 on E: $(P_1 + P_2) + P_3) = P_1 + (P_2 + P_3)$ (Associativity)*

Therefore the points of E with their addition operator form an abelian group, where $\mathcal{O}$ is the identity element.

Proof. **Commutativity:** For any P_1 and P_2 on E the line through P_1 and P_2 is the same as the line through P_2 and P_1, so they yield the same P_3. Hence $P_1 + P_2 = P_2 + P_1$.

Identity: This property holds by the definition of $\mathcal{O}$

Inverses: For any point P on E let $-P$ be the point obtained when P is reflected across the x-axis. Then the line through P and $-P$ is vertical and follows

Case 2, so that $P + (-P) = \mathcal{O}$. If E is in the Weierstrass form (2.2) and if $P = (x, y)$, then $-P = (x, -y)$

Associativity: The proof of this property is very long and not necessary for the remainder of this paper. Therefore for a proof of associativity see [14]

□

From the above proof we can see that if three points all lie on the same line then one point is the inverse of the sum of the other two. In the earlier example we had $P + Q = R$ where $R = -S$ so that P, R, S were all on one line then $P + Q + S = \mathcal{O}$.

2.3 The j-invariant

We have shown it is sometimes possible to transform one elliptic curve into another, so that we know two curves are isomorphic. However, if we are just given the two curves, then seeing that they are isomorphic might not be so easy. In that case calculating the j-invariants of the curves is helpful.

Definition 6. *(j-invariant) The* j-invariant *of an elliptic curve E given in the short Weierstrass form is*

$$j(E) = 1728\frac{4A^3}{4A^3 + 27B^2}$$

Notice that the denominator of the j-invariant was assumed to be nonzero, thus the j-invariant is defined for all of our curves. Now consider the change of variables $x_1 = \mu^2 x$ and $y_1 = \mu^3 y$, (given in [14]) for the curve $y^2 = x^3 + Ax + B$ where $\mu \in \overline{K}\backslash\{0\}$. Then

$$\begin{aligned} y^2 &= x^3 + Ax + B \\ \left(\mu^{-3}y_1\right)^2 &= \left(\mu^{-2}x_1\right)^3 + A\mu^{-2}x_1 + B \\ \mu^{-6}y_1^2 &= \mu^{-6}x_1^3 + A\mu^{-2}x_1 + B \\ y_1^2 &= x_1^3 + A\mu^4 x_1 + \mu^6 B \\ y_1^2 &= x_1^3 + A_1 x_1 + B_1. \end{aligned} \tag{2.4}$$

For $A_1 = \mu^4 A$ and $B_1 = \mu^6 B$. Thus we have two equivalent equations whose change of variables did not affect the j-invariant.

Theorem 2. *Two elliptic curves given in the Weierstrass form are isomorphic to each other in $\overline{K}$ if and only if they have the same j- invariant.*

Proof. Let $E_1 : y^2 = x^3 + A_1 x + B_1$ and $E_2 : y^2 = x^3 + A_2 x + B_2$ be isomorphic to each other such that E_2 can be transformed from E_1 in the same matter that (2.4) was obtained. Then $A_2 = \mu^4 A_1$ and $B_2 = \mu^6 B_1$ and by calculating the j-invariants we get

$$\begin{aligned} j(E_2) &= 1728\frac{4A_2^3}{4A_2^3 + 27B_2^2} \\ &= 1728\frac{4(\mu^4 A_1)^3}{4(\mu^4 A_1)^3 + 27(\mu^6 B_1)^2} \\ &= 1728\frac{4\mu^{12} A_1^3}{4\mu^{12} A_1^3 + 27\mu^{12} B_1^2} \\ &= 1728\frac{4A_1^3}{4A_1^3 + 27B_1^2} \\ &= j(E_1) \end{aligned}$$

Therefore the two isomorphic curves have the same j-invariant. For the converse, let $j(E_1) = j(E_2)$. First consider the case where $A_1 = 0$ then $j(E_1) = 0 = j(E_2)$ so that $A_2 = 0$. Since $4A^3 + 27B^2 \neq 0$, $B_1, B_2 \neq 0$. Then chose a μ such that $B_2 = \mu^6 B_1$, then E_1 is isomorphic to E_2. Now suppose that $A_1 \neq 0$, then choose a μ such that $A_2 = \mu^4 A_1$, by substituting $A_2\mu^{-4} = A_1$ in the j-invariant formula we get

$$\begin{aligned} j(E_1) &= j(E_2) \\ 1728\frac{4A_1^3}{4A_1^3 + 27B_1^2} &= 1728\frac{4A_2^3}{4A_2^3 + 27B_2^2} \\ \frac{4\mu^{-12}A_2^3}{4\mu^{-12}A_2^3 + 27B_1^2} &= \frac{4A_2^3}{4A_2^3 + 27B_2^2} \\ \frac{4A_2^3}{4A_2^3 + 27\mu^{-12}B_1^2} &= \frac{4A_2^3}{4A_2^3 + 27B_2^2}. \end{aligned}$$

This implies that $B_2^2 = (\mu^6 B_1)^2$, so $B_2 = \pm\mu^6 B_1$. If $B_2 = \mu^6 B_1$ then we are done. If $B_2 = -\mu^6 B_1$, then instead of choosing μ chose $i\mu$, where $i^2 = -1$. Then $A_2 = \mu^4 A_1$ is left unchanged but we now obtain $B_2 = -\mu^6 B_1$, and once again E_1 is isomorphic to E_2.

□

This argument above only holds if we are considering two curves over an algebraically closed field, since we are taking roots in the field. Over a general field it is possible for two curve to have the same j-invariant but that there does not exist a transformation between the two in that field.

Definition 7. *(Twist) For two curves E_1 and E_2 both over K, where $E_1 \neq E_2$, if $j(E_1) = j(E_2)$ then they are twist of each other.*

In particular, for any nonzero $c \in K$, the *c-twist* of an elliptic curve in the short Weierstrass form is given by

$$y^2 = x^3 + Ac^2x + Bc^3 \tag{2.5}$$

By a simple calculation we can see that (2.5) has the same j-invariant as the curve (2.2) does. Although we will not be calculating the j-invariant for the other forms of elliptic curves, we will discuss the transformations between them.

CHAPTER III

ELLIPTIC CURVES OVER FINITE FIELDS

Elliptic curves can be defined over finite fields, such as the field $\mathbb{Z}_p$ where p is a prime. Thus we have a finite abelian group and all the usual things that are associated with that. Take for example the curve $E(\mathbb{Z}_{13}) : y^2 = x^3 + 3x + 5$. It contains the point $Q = (4 : 9 : 1)$ and using the addition formulas we have that $Q + Q = (4 : 4 : 1)$ and that $Q + Q + Q = (0 : 1 : 0)$. So if we add Q to itself three times we end up at the point of infinity. Note that all the addition calculations are being done modulo p.

Definition 8. *(Order of a point) If P is a point on a curve, then $n*P = \overbrace{P + P + \ldots + P}^{n \text{ times}}$. The* order of a point *is the smallest positive integer, n, such that $n * P = \mathcal{O}$.*

So in the previous example we have that the $order(Q) = 3$. When the order of a point is two, it leads to an interesting fact.

Proposition 2. *Let E be an elliptic curve in the Weierstrass form, and let P be a point on E of order 2, then $f(x) = x^3 + Ax + B$ has at least one root in K.*

Proof. Let $P = (x_1, y_1)$. Since the order of P is 2, $P + P = \mathcal{O}$ and so $P = -P$. In the Weierstrass form that means $-P = (x_1, -y_1)$, so that $(x_1, y_1) = (x_1, -y_1)$ and implies $y_1 = 0$. Therefore when $x = x_1$, $x^3 + Ax + B = 0$, so x_1 is a root of $f(x)$.

□

If K is a finite field then there can only exist a finite number of points on a curve given over K. Let $\mathbb{F}_q$ represent a finite field of q elements, where $q = p^k$ for a prime p, and a natural number k.

Definition 9. *(Order of an Elliptic Curve) For a finite field $\mathbb{F}_q$, the order of a elliptic curve $E(\mathbb{F}_q)$ is number of points on E in $\mathbb{F}_q$, denoted as $\#E(\mathbb{F}_q)$.*

So in the example of $E(\mathbb{Z}_{13}) : y^2 = x^3 + 3x + 5$, $\#E(\mathbb{Z}_{13}) = 9$. Notice that the order of Q divides the order of its curve, this is always true, since the curve now form a finite abelian group. In general determining the exact order of a curve is hard, especially as the size of the field increases. However we do have bounds on what those orders can be, due to Hasse. (For a proof see [14].)

Theorem 3. *(Hasse) For an elliptic curve $E(\mathbb{F}_q)$*

$$\#E(\mathbb{F}_q) = q + 1 - t_q \qquad \textit{where} \quad |t_q| \leq 2\sqrt{q}.$$

The quantity t_q is called the *trace of Frobenius* for $E(\mathbb{F}_q)$. As we will soon see, the order of a curve is an extremely useful property of elliptic curves.

CHAPTER IV

THE ELLIPTIC CURVE FACTORING METHOD

Now consider the case when an "elliptic curve" is defined over $\mathbb{Z}_N$, where N is a composite of two or more primes. Since $\mathbb{Z}_N$ is clearly not a field, we cannot expect do elliptic curve arithmetic on the points that satisfy the cubic over $\mathbb{Z}_N$ as we did for the fields above. We no longer have a group, since for some points P and Q on $E(\mathbb{Z}_N)$, the sum $P + Q$ might not exist. The failure of this sum to exist is due to the addition method. In section (2.2) the slope, m is calculated every time an addition is preformed and in doing so an element of $\mathbb{Z}_N$ must be inverted. This works fine for non-zero elements in a field, since they all will have inverses. However in $\mathbb{Z}_N$, there exist elements that do not have inverses, namely those elements that are not relatively prime to N. Now while trying to add two points on our pseudo curve using the same addition methods as before, if such an element, g, happened to be in the denominator of the slope, then that sum would not exist, because $GCD(N, g) \neq 1$. Well, if we wanted to factor that N, then this would actually be a good thing, since we would have found a factor of N. This is the idea behind Lenstra's *Elliptic Curve Factoring Method*, often referred to as ECM.

It was in 1985 that H. Lenstra first introduced the ECM and since then there have been many of contributions and improvements to it. Lenstra's original algorithm is now known as phase 1. The introduction of phase 2 is due to Montgomery and Brent, see [10]. The ECM is currently the third fastest known factorization algorithm after the Quadratic and Number Field Sieves, and is much easier to implement. It is generally used to factor out medium size factors, those around 20 to

30 digits, although larger factors have been found using the ECM.

4.1 Phase 1

The idea Lenstra had for Phase 1 is based upon *Pollard's* $p-1$ *algorithm.* This algorthim uses the fact that if p is a prime and a an integer, where p does not divide a, then $a^{p-1} \equiv 1 \bmod p$ and thus p divides $a^{p-1}-1$ (Fermat's Little Theorem.) Suppose, by some means, we have an integer K, which $p-1$ divides so that $K = j(p-1)$. Then for some a,

$$a^K = a^{j(p-1)} = a^{(p-1)^j} \equiv 1^j \equiv 1 \quad mod\, p. \tag{4.1}$$

Then the $GCD(a^K - 1, N) > 1$, and we maybe have found a factor. It could be N, but then choosing a different a might still yield a factor. (See [7] for more on Pollard's method.) A similar scheme to (4.1) is what makes the ECM work. First let's consider the algorithm for Phase 1, then we will investigate why it works.

Algorithm 1. *(ECM Phase 1)*

To factor an integer N*:*

1. *Choose some bound* B_1.
2. *Define* $L = \prod p_i^{a_i}$ *for all primes,* $p_i < B_1$ *where* $a_i = \lfloor \frac{\log B_1}{\log p_i} \rfloor$.
3. *For* i *in some chosen interval do*
   - *i. Chose a point* $P_i = (s_i, t_i)$ *and an* a_i*, all in* $\mathbb{Z}_N$.
   - *ii. Set* $b_i = t_i^2 - s_i^3 - a_i s_i$. *//This defines the curve* $E_i : y^2 = x^3 + a_i x + b_i$.
   - *iii. If* $4a_i^3 + 27b_i^2 = 0$*, then go back to beginning of loop, else*
   - *iv. Compute* $Q_i = L * P_i$ *//Using the formulas in section 2.2*

v. *If while trying to computer Q_i it fails, then it is because some number, g, was not invertible.*

Compute $GCD(N, g) = q$, then return q.

4. *If Q is successfully computed, then increment i and end loop.*
5. *If i's interval has been exhausted return Fail or got to phase 2.*

Here all the arithmetic is being done modulo N, which is not true elliptic curve arithmetic. However, underneath it there are actual group operations being done on the curve $E(\mathbb{F}_q)$, obtained by reducing everything in $E(N)$ modulo q, for q a prime factor of N. The hope is that the order of $E(\mathbb{F}_q)$ divides L so that in computing $L * P_i$ we must step through $\mathcal{O}$ of $E(\mathbb{F}_q)$. This will result in an non invertible element being in the denominator of the slope.

Definition 10. *(Smooth) An integer N is* B-smooth*, if all of its prime factors are less than or equal to the bound B.*

So we hope that the $\#E(\mathbb{F}_q)$ is B-smooth for some prime dividing N which wil lead to $\#E(F_q) \mid L$. Notice that Pollard's method relies on the same hope, that the group order of $\mathbb{Z}_{N-1}$ is smooth. The advantage of the ECM over Pollard's method is that if one curve's order is not smooth, then we have plenty more curves to work with. Where as in Pollard's method we are stuck with just the one group.

4.2 Phase 2

If phase 1 does not yield a factor then we were unable to find a curve whose order was B-smooth, but perhaps we were close. In phase 2 we are considering that maybe $\#E\mathbb{F}_q)$ is B-smooth, except for one prime, p, that exceeds B_1. In other words

$$\#E(\mathbb{F}_q) = p * L.$$

The algorithm of Phase 2 uses projective coordinates for the points. Though we have not yet discussed addition with projective coordinates, two addition formulas will be given in the remaining chapters, along with how to implement them into the ECM. See 5.2 and 6.1. For now we will just discuss the idea of the algorithm. The following version is similar to the one given in [8].

Algorithm 2. *(ECM Phase 2) Note Q is the point returned at the end of Phase 1 before the loop repeats*

1. *Chose another bound* $B_2 > B_1$.
2. *Set t=1*
3. *For every prime, p such that* $B_2 > p > B_1$ *do*
   i. *Calculate* $pQ = (x_{pQ} : y_{pQ} : z_{pQ})$
   ii. $t = t \cdot z_{pQ}$ *mod N.*
4. *Calculate* $GCD(t, N) = d$
5. *If* $d > 1$ *return d.*
6. *Else return Fail.*

Recall that points in the projective plane whose z coordinate is zero corresponds to a point of infinity. So for some point $P = (x_p : y_p : z_p)$ and q a factor of N, if $z_p \equiv 0 \bmod q$ but not mod N, then the $GCD(N, z_p) = q$. Furthermore any multiplication to z_p once this occurred would have no effect on the GCD. Therefore we save operations by simply multiplying all the z coordinates together and taking the GCD only once at the end of the phase.

The importance of the other two curves in the ECM, is their addition methods. Since most of the time the ECM spends is on calculating this large multiple of

a point, we would like to find faster ways to add points together. The Montgomery and Edwards curve can speed up the addition by use of projective coordinates.

CHAPTER V

MONTGOMERY CURVES

First introduced in 1987 by Peter Montgomery, Montgomery curves appear extensively in his dissertation, see [9]. These types of curves can greatly increase the speed of the ECM, since we can define the addition method in such a way that we do not need to compute a GCD every time we calculate the sum of two points. *Montgomery curves* are elliptic curves in the form

$$BY^2 = X^3 + AX^2 + X \tag{5.1}$$

where $A, B \in K$, $B \neq 0$ and $A \neq \pm 2$. (For if $A = 2$ then a double roots is obtain, namely $x = -1$.) All curves in Montgomery form have the point $(0, 0)$ which is of order 2. Not every curve in the Weierstrass form can be transformed into a Montgomery curve, since there exist curves that do not have any points of order 2 (while staying in K, and not going into $\overline{K}$.) However, in [11] conditions are given to determine if a curve is transformable between the Weierstrass and Montgomery forms .

Theorem 4. *An elliptic curve in the Weierstrass form $E_W : y^2 = x^3 + ax + b$ is transformable to the Montgomery form $E_M : BY^2 = X^3 + AX^2 + X$ if and only if the following conditions are satisfied.* [11]

1. *For the equation $f(x) = x^3 + ax + b$, there exist an $\alpha \in K$ such that $f(\alpha) = 0$.*
2. *The quantity $(3\alpha^2 + a)$ is a quadratic residue in K.*

Proof. First, assume that E_W satisfies the conditions and let s, A, and B be as follows

$$s^2 = (3\alpha^2 + a)^{-1} \qquad A = 3\alpha s \qquad B = s.$$

Then for any point (x, y) on E_W the mapping $(x, y) \to (s(x - \alpha), sy)$ gives us the transformation from E_W to E_M, in the following way. First, start with E_M and substitute in $X = s(x - \alpha), Y = sy$, A, and B as they were defined above, so that

$$\begin{aligned}
BY^2 &= X^3 + AX^2 + X \\
s(sy)^2 &= (s(x-\alpha))^3 + (3\alpha s)(s(x-\alpha))^2 + (s(x-\alpha)) \\
s^3y^2 &= s^3(x^3 - 3x^2\alpha + 3x\alpha^2 - \alpha^3) + 3\alpha s^3(x^2 - 2\alpha x + \alpha^2) + s(x-\alpha) \\
y^2 &= (x^3 - 3x^2\alpha + 3x\alpha^2 - \alpha^3) + 3\alpha(x^2 - 2\alpha x + \alpha^2) + s^{-2}(x-\alpha) \\
&= x^3 - 3x^2\alpha + 3x\alpha^2 - \alpha^3 + 3\alpha(x^2 - 2\alpha x + \alpha^2) + (3\alpha^2 + a)(x-\alpha) \\
&= x^3 + (-3\alpha + 3\alpha)x^2 + (3\alpha^2 - 6\alpha^2 + 3\alpha^2 + a)x + (-\alpha^3 + 3\alpha^3 - 3\alpha^3 - a\alpha) \\
&= x^3 + ax + (-\alpha^3 - a\alpha).
\end{aligned}$$

Recall that $\alpha^3 + a\alpha + b = 0$, which implies that $a\alpha = -\alpha^3 - b$ so we have

$$y^2 = x^3 + ax + (-\alpha^3 + \alpha^3 + b)$$

$$y^2 = x^3 + ax + b.$$

Thus we have ended with E_W.

Now assume that E_W is transformable to E_M, then there must exist a point in E_W that has order 2. Therefore, by Proposition 2 and the fact that $x^3+ax+b = 0$ has a root, condition 1 has been satisfied. An isomorphic mapping from E_W to E_M is given that $(x, y) \to (s(x - \alpha'), t(y - \beta'))$ for some $s, t, \alpha', \beta' \in K$ where $s, t \neq 0$. Now α is a root on E_W, so it is of order 2 and therefore must be mapped to a point

of order 2 on E_M and since every curve in Montgomery form has the point $(0,0)$ of order 2, $(\alpha,0) \to (0,0)$. Which makes $\alpha' = \alpha$ and $\beta' = 0$. So the mapping from E_W to E_M is given by $(s(x-\alpha), ty)$. This point is on E_M, and by substituting it in we obtain

$$Bt^2y^2 = s^3(x-\alpha)^3 + As^2(x-\alpha)^2 + s(x-\alpha).$$

Recall that $f(x) = x^3 + ax + b = y^2$, then

$$Bt^2(x^3 + ax + b) = s^3(x-\alpha)^3 + As^2(x-\alpha)^2 + s(x-\alpha).$$

Now by comparing the x^3 coefficients, we obtain $Bt^2 = s^3$ so for

$$Bt^2 f(x) = s^3(x-\alpha)^3 + As^2(x-\alpha)^2 + s(x-\alpha)$$

$$s^3 f(x) = s^3(x-\alpha)^3 + As^2(x-\alpha)^2 + s(x-\alpha)$$

$$s^2 f(x) = s^2(x-\alpha)^3 + As(x-\alpha)^2 + (x-\alpha).$$

Now taking the derivative with respect to x, then evaluating at $x = \alpha$ we obtain

$$s^2 f'(x) = 3s^2(x-\alpha)^2 + 2As(x-\alpha) + 1$$

$$s^2 f'(\alpha) = 1$$

$$f'(\alpha) = 3\alpha^2 + a = \frac{1}{s^2}.$$

Therefore $3\alpha^2 + a$ is quadratic residue in K, and thus condition 2 has been satisfied.

□

Suppose we have the curve $E_W(\mathbb{Z}_{11})$ defined by $y^2 = x^3 + 5x + 2$. Then condition 1 holds since in $\mathbb{Z}_{11}$, $\alpha = 3$ is a root of $x^3 + 5x + 2 = 0$. Condition 2 is also satisfied since $(3\alpha^2 + a) = 3(3)^2 + 5 \equiv 9 \equiv 3^2 \bmod 11$. Using the change of variables,

$$A = 3\alpha s, \qquad B = s, \qquad \text{where} \qquad s = \frac{1}{3\alpha^2 + a}$$

we have $s = \frac{1}{9} \equiv 5 \bmod 11$, which means that $A = 1$, and $B = 5$ therefore

$$E_M(\mathbb{Z}_{11}) : 5y^2 = x^3 + x^2 + x$$

5.1 Montgomery Addition

For our factoring purposes, we are interested in computing multiples of points. It turns out that we can speed up computations in the case of computing multiples. Therefore, our discussion of addition on Montgomery curves is restricted to multiples of a fixed point. Also we will consider points to be given in their projective coordinates, $(x : y : z)$, and for this we will use $BY^2Z = X^3 + AX^2Z + XZ^2$, the homogeneous form of E_M. We can see that here $\mathcal{O}$ is also $(0 : 1 : 0)$. We will see that in this form addition is an inversion free process, and that we can ignore the y-coordinate. Thus, points are simply written as $(x :: z)$. In exchange for not having the y-value, we now will have to know the difference of two points to be able to find their sum.

Notation: If $P = (x :: z)$ is a point on E_M, then for adding P to itself n times we write

$$n * P = P_n = (x_n :: z_n).$$

Addition: For points P_m and P_n on E_M for $m \neq n$, $\quad P_{m+n} = (x_{m+n} :: z_{m+n})$

where

$$x_{m+n} = z_{m-n}((x_m - z_m)(x_n + z_n) + (x_m + z_m)(x_n - z_n))^2 \tag{5.2}$$
$$z_{m+n} = x_{m-n}((x_m - z_m)(x_n + z_n) - (x_m + z_m)(x_n - z_n))^2.$$

Doubling: If $m = n$, then $P_{2n} = (x_{2n} :: z_{2n})$ is given by

$$\begin{aligned} x_{2n} &= (x_n + z_n)^2(x_n - z_n)^2 \\ z_{2n} &= (4x_n z_n)((x_n - z_n))^2 + (A + 2/4)(4x_n z_n) \end{aligned} \tag{5.3}$$

where

$$4x_n z_n = (x_n + z_n)^2 - (x_n - z_n)^2.$$

Consider the curve $2y^2 = x^3 + 6x^2 + x$ which has the point $P = (1, 2)$ so we represent it as $P_1 = (1 :: 1)$ then P_2 is calculated by using the doubling formula.

$$\begin{aligned} 4x_1 z_1 &= (x_1 + z_1)^2 - (x_1 - z_1)^2 = (1 + 1)^2 - (1 - 1)^2 = 4 \\ x_2 &= (x_1 + z_1)^2(x_1 - z_1)^2 = (1 + 1)^2(1 - 1)^2 = 0 \\ z_2 &= (4x_1 z_1)((x_1 - z_1)^2 + ((A + 2)/4)(4x_1 z_1)) \\ &= 4[(1 - 1)^2 + (2)(4)] = 32 \end{aligned}$$

So $P_2 = (0 :: 32)$.

We could now find $P_1 + P_2 = P_3$, since we know the difference of P_1 and P_2. If we wanted to compute a large multiple of a point P_m where $m = r + s$, then in addition to having to know r and s, we would need to know P_{r-s}. Therefore, its seems that in the process of computing a multiple we would have to remember the smaller multiples along the way. However in [9] an addition ladder is given so that we can compute a larger multiple quickly, while only knowing three points: the two points we want to add, (P_r, P_s) and the original point we first started with, P. (The

version below is adapted from [6].) Given some curve E_M that has points P_m, P_n and D (all multiples of a single point), let Madd(P_m, P_n, D) be the process defined by (5.2) so it gives us the sum P_{m+n}, where $D = P_{m-n}$. Also, let Mdbl (P_n) be the process defined by (5.3), so it gives us the point P_{2n}.

Algorithm 3. *(Montgomery Addition Ladder)*

Given a point P and an integer $n > 2$, to calculate P_n :

1. *Represent n as a binary sequence of bits $\{n_{B-1}, \ldots, n_0\}$, where $n = \sum_{i=0}^{B-1} 2^i n_i$*
2. *Set $V = P$ and $W = Mdbl(P)$.*
3. *Loop over the bits of n, starting with n_{B-2} down to n_0.*
   i. *If $n_i = 1$ then*
   $$V = Madd(W, V, P)$$
   $$W = Mdbl(W)$$
   ii. *Else*
   $$W = Madd(V, W, P)$$
   $$V = Mdbl(V)$$
4. *End loop.*
5. *Return V*

This is how the large multiples in the ECM are calculated more efficiently. In algorithm 1 we get a factor of N if while adding two points we comes across a element of $\mathbb{Z}_N$ that in not invertible, but Montgomery addition is inversion free. As before with phase 2, (algorithm2) we are concerned about the z coordinate of our point, however in [12] the following fact is given.

Proposition 3. *Let $q = a + b$ with a and b relatively prime and let n be an integer with t as a prime factor. For a point $Q = (X_Q :: Z_Q)$ on E_M, let $A = aQ$ and $B = bQ$ so that $qQ = A + B$. Then for the $GCD(Z_Q, n) = 1$,*

$$Z_{qQ} \equiv 0 \ mod\ t \ \textit{if and only if}\ X_A Z_B - Z_A X_B \equiv 0 \ mod\ t.$$

Proof. If $t \mid (X_A Z_B - Z_A X_B)$ then also $t \mid (X_{A-B}(X_A Z_B - Z_A X_B)^2)$, by (5.2) we know $Z_{qQ} = Z_{A+B} = X_{A-B}(X_A Z_B - Z_A X_B)^2$. Thus $t \mid Z_{qQ}$, so $Z_{qQ} \equiv 0 \bmod t$.

Conversely if $Z_{qQ} \equiv 0 \bmod t$ that implies that $qQ \equiv \mathcal{O}$ over the field of t elements. Therefore $qQ \equiv \mathcal{O} \equiv A + B$ so that $A = -B \bmod t$. Then, either $A = B = 0$ which case $Q = 0 \bmod t$, so $t \mid Z_Q$ which is a contradiction to the assumption that $GCD(Z_Q, n) = 1$, otherwise

$$\frac{X_A}{Z_A} \equiv \frac{X_B}{Z_B} \bmod \text{t}$$

since the only the y value differs in inverses. Then we have

$$X_A Z_B \equiv Z_A X_B \bmod \text{t}$$

$$X_A Z_B - Z_A X_B \equiv 0 \bmod \text{t}$$

Hence $t \mid (X_A Z_B - Z_A X_B)$.

□

So we do not need to actually calculate the point qQ, we just need to know the coordinates of the points we are adding together to get it. This enhancement will be used in phase 2 of the ECM.

5.2 Implementation of ECM

The following modifications can be made to algorithms 1 and 2 so that they are suitable for Montgomery curves. Here we have followed the implementation in [13].

Notation: If we want to calculate the point P added to itself $n+10$ times we write $[n+10]P$.

Algorithm 4. *(ECM for Inverse-free Addition)*

To factor an integer N:

Phase 1

I. Choose a bound B_1. *//It must be even.*

II. Set $B_2 = 100B_1$.

III. Define $L = \prod p_i^{a_i}$ *for all primes,* $p_i < B_1$ *where* $a_i = \lfloor \frac{\log B_1}{\log p_i} \rfloor$.

IV. For i in some chosen interval do

1. *Chose a point* $P = (s :: u)$ *where* $s, u \in \mathbb{Z}_N - \{0\}$.
2. *Set* $A = \frac{u - s^3 + su^2}{s^2 u}$. *//Defines the curve* $E_M : y^2 = x^3 + Ax^2 + x$.
3. *If* $A = \pm 2$ *then go to beginning of loop, else*
4. *Compute* $Q = (x :: z) = L * P$. *//Using algorithm (3).*
5. *Let* $t = GCD(N, z)$.
6. *If* $t > 1$ *then return t, else*

Phase 2

7. *Set* $S_1 = Mdbl(Q)$ *and* $S_2 = Mdbl(S_1)$.
8. *For d in* $[3, 100]$, *set* $S_d = Madd(S_{d-1}, S_1, S_{d-2})$ *mod N.*
9. *Set* $t = 1$, $B = B_1 - 1$ *and* $r = B$.
10. *Compute* $V = [B - 200]Q$ *and* $R = [B]Q$. *//Using algorithm (3)*
11. *For* $r < B_2$ *do*

    a. *For prime q in* $[r + 2, r + 200]$ *do*

i. *Set* $\delta = \frac{q-r}{2}$.

ii. *Set* $t = t * (X_R Z_{S_\delta} - Z_R X_{S_\delta})$.

b. *Set* $V = R$ *and* $R = Madd(R, S_{100}, V)$.

c. *Increment* r *by 200.*

12. *Calculate* $d = GCD(t, N)$.

13. *If* $d > 1$ *return* d.

14. *Else increment* i *by 1.*

15. *If* i*'s interval has been exhausted return Fail.*

If p is a prime factor of N, the point P in phase 1 reduces to a point on a genuine elliptic curve, $E(\mathbb{F}_p)$. If the order $E(\mathbb{F}_p)$ is B_1 smooth then $Q = L * P$ is equivalent to $\mathcal{O}$ mod p, moreover $z_Q \equiv 0 \bmod p$. Thus p divides z_Q and N, so we have found a factor. If phase 1 fails then the hope in phase 2 is that we missed the smoothness of the order of $E(\mathbb{F}_p)$ by just one prime. Therefore we only compute $q * Q$ for every prime $B_1 < q < B_2$, instead of the product of all the primes as we did in phase 1. Since, in phase 2, we no longer add points a single multiple at a time, we are required to compute and store multiples of Q. This is what the S_d are in the algorithm above, for every $d \in [1, 100]$ we compute and store $S_d = [2d]Q$. Then we calculate the distance to the next prime, the δ tells us which S multiple we are to use. Proposition (3) tells us we only need the two points we would add to get qQ and not the actual point. So with the points S_q and R we can test if q is that outlying prime.

The success of the ECM relies on choosing a bound, B so that the order of $E(F_p)$ is B-smooth. If we want to find a prime factor, q, of N then knowing the probability of if q is B-smooth is something to consider.

Definition 11. *($\psi(x,y)$) Let x and y be integers then*

$$\psi(x,y) = \#\{1 \leq n \leq x \mid n \text{ is } y\text{-smooth}\}$$

The following theorem is given in [13].

Theorem 5. *(Dickman) For each real number $u > 0$, there exist a real number $\rho(u) > 0$ such that*

$$\psi(x, x^{\frac{1}{u}}) \cong \rho(u)x$$

So we have that $\rho(u)$ is the probability that a number less than or equal to x is $x^{\frac{1}{u}}$ smooth. The actual calculation of $\rho(u)$ is a very complex function. However it can be approximated by

$$\rho(u) \approx u^{-u}.$$

Therefore if we want to know the probability that a number less than or equal to n is B-smooth we can calculate,

$$\frac{\log n}{\log B} = a$$

$$\rho(a) \approx a^{-a}$$

Say we want to extract an n-bit prime factor, q from N and we want to choose a bound, $B = 2^k$ in such a way that half of the number less than or equal to q are B-smooth, in other words $\rho(\frac{n}{k}) = \frac{1}{2}$. Define $u(e)$ to be a number such that

$\rho(u(e)) = e$. Therefore

$$\begin{aligned}
\rho(\frac{\log q}{\log B}) &= \frac{1}{2} \\
\frac{\log q}{\log B} &= u(\frac{1}{2}) \\
\log B &= \frac{\log q}{u(\frac{1}{2})} \\
\log B &= \frac{n}{u(\frac{1}{2})} \\
k &= \frac{n}{u(\frac{1}{2})}.
\end{aligned}$$

Thus if we choose our k in this manner about half of the time a number $j \leq q$ would be B-smooth. Now if $j = \#E(\mathbb{F}_q)$, then the ECM would produce the factor q. To compute $u(e)$ see appendix A.

CHAPTER VI

EDWARDS CURVES

The last family of curve we will discuss were introduced in 2007 by Harold Edwards in his paper [5]. These curves have naturally become known as Edward's curves. All elliptic curves over a field of characteristic not 2 can be transformed to an Edward's curve, sometimes at the expense of going into an extension of the original field. The most general form of an *Edwards curve* is given by

$$E_E : x^2 + y^2 = c^2(1 + dx^2y^2)$$

where $cd(1 - dc^4) \neq 0$. Often there is special consideration given for when $c^2 = 1$.

Proposition 4. *If there exists an element $d_1 \in K$ such that $d_1 = dc^4$ for some $c, d \in K$ then*

$$x^2 + y^2 = c^2(1 + dx^2y^2)$$

can be transformed to

$$x_1^2 + y_1^2 = 1 + d_1x_1^2y_1^2.$$

Proof. Let $x = cx_1$ and $y = cy_1$ then by simple substitution we have

$$\begin{aligned}
x^2 + y^2 &= c^2(1 + dx^2y^2) \\
(cx_1)^2 + (cy_1)^2 &= c^2(1 + d(cx_1)^2(cy_1)^2) \\
c^2x_1^2 + c^2y_1^2 &= c^2(1 + dc^4x_1^2y_1^2) \\
x_1^2 + y_1^2 &= 1 + dc^4x_1^2y_1^2 \\
x_1^2 + y_1^2 &= 1 + d_1x_1^2y_1^2
\end{aligned}$$

□

Therefore we will only consider Edwards curves where $c = 1$ with the added exemption that d is not a square.

6.1 Edwards Addition

Just as before, we have a way to adding two points on an Edwards curve. However here the identity element is different, as well as the inverses. Now the identity is given by $(0, 1)$ and the inverse of a point (x, y) is $(-x, y)$. This addition can be defined in the affine coordinates or can be expanded to the projective coordinate where, as in Montgomery curves, the addition process is inversion free.

Affine Coordinates:

For two points $P = (x_p, y_p)$ and $Q = (x_q, y_q)$ on E_E define $P + Q = R = (x_r, y_r)$ for

$$\begin{aligned} x_r &= \frac{x_p y_q + x_q y_p}{(1 + d x_p x_q y_p y_q)} \\ y_r &= \frac{y_p y_q - x_p x_q}{(1 + d x_p x_q y_p y_q)}, \end{aligned}$$

where $d x_p x_q y_p y_q \neq \pm 1$. If $Q = -P$ then $Q = (-x, y)$ and

$$\begin{aligned} P + (-P) &= \left(\frac{xy - xy}{(1 - dxxyy)}, \frac{yy + xx}{(1 - dxxyy)} \right) \\ &= (0, \frac{y^2 + x^2}{1 - dx^2y^2}) \\ &= (0, \frac{y^2 + x^2}{y^2 + x^2}) \\ &= (0, 1). \end{aligned}$$

Thus we get the identity element as expected.

Projective Coordinates:

Before we can define the addition for the projective coordinates, we first we have to put the equation in its homogeneous form,

$$(X^2 + Y^2)Z^2 = Z^4 + dX^2Y^2.$$

Now the identity is represented as $(0 : 1 : 1)$, and the inverse of a point $(X : Y : Z)$ is $(-X : Y : Z)$. These projective points correspond the points $(X/Z, Y/Z)$ in the affine space. For two points $P = (X_p : Y_p : Z_P)$ and $Q = (X_q : Y_q : Z_p)$ on E_E, define $P + Q = R = (X_r : Y_r : Z_r)$ as follows.

$$\begin{aligned} X_r &= (Z_p^3 Z_q^3 - dX_pX_qY_pY_qZ_pZ_q)((X_p + Y_p)(X_q + Y_q) - (X_pX_q + Y_pY_q)) \\ Y_r &= (Z_p^3 Z_q^3 + dX_pX_qY_pY_qZ_pZ_q)(Y_pY_q - X_pX_q) \\ Z_r &= Z_p^4 Z_q^4 - d^2 X_p^2 X_q^2 Y_p^2 Y_q^2 \end{aligned} \tag{6.1}$$

When actually using this addition formulas in the ECM, it is given in different parameters. Notice the repetitiveness of the formulas, if we define pieces of them then the overall calculation will use less operations, thus making the addition faster. [3] uses the following formulas.

$$A = Z_pZ_q \qquad B = A^2 \qquad C = X_pX_q \qquad D = Y_pY_q$$

$$D = Y_pY_q \qquad E = dCD \qquad F = B - E \qquad G = B + E$$

$$\begin{aligned} X_r &= AF((X_p + Y_p)(X_q + Y_q) - C - D) \\ Y_r &= AG(D - C) \\ Z_r &= FG \end{aligned} \tag{6.2}$$

To see if this addition method (6.2), is actually faster we have constructed a program that measures the time that each algorithm spent on the computation of a large multiple. (see Appendix A for the code) In addition to comparing the two different parameterizations of Edwards curves, we have also included Montgomery curves. To make this comparison as fair as possible we started with an Edwards curve and constructed an isomorphic Montgomery curve, via the parameters given in [1]. After runing this program several thousand times, we have seen that the Edwards addition given by algorithm (6.2) is consistently faster.

The ECM algorithm (4), given for Montgomery curves can be easily adjusted to accommodate Edward curves. The only change to (4) would be to the addition algorithm so that it has the parameters of (6.2). We have determined that this is the fastest addition method discussed here, and therefore it would be our best choice for use in the ECM, since it would make the overall algorithm faster as well.

CHAPTER VII

CONCLUSION

While all elliptic curves can be write in the Weierstrass form, if we consider other forms we can improve the speed to the ECM. The implementation of Montgomery curves into the ECM was the first step. The use of projective coordinates eliminated the need of computing the GCD in its addition method. The newer form of curves, the Edwards curves, provide an even faster addition method, which we have been able to demonstrate through our program that compares computation times. Although we have only discussed which addition method is quicker, there are many more improvements that can be done to speed up the ECM. We hope in future endeavors that we will be able to explore those ideas, as well as to go deeper in to elliptic curve theory, in particular the geometric motivation of the Edwards addition laws.

BIBLIOGRAPHY

[1] Bernstein, Daniel, Peter Birkner, Christiane Peters, Tanja Lange, and Marc Joye, "Twisted Edwards Curves," 2008, id. c798703ae3ecfdc3751 12f19dd0787e4.

[2] Bernstein, Daniel, Peter Birkner, Tanja Lange and Christiane Peters, "ECM using Edwards curves" 2008, id. cb39208064693232e4751ec8f3494c43.

[3] Bernstein, Daniel and Tanja Lange, "Faster addition and doubling on elliptic curves," 2007 id. 95616567a6ba20f575c5f25e7cebaf83.

[4] Cannon, J. J. W. Bosma (Eds.) Handbook of Magma Functions, Edition V2.14-14 (STUDENT), 2006, 4350 pages.

[5] Edwards, Harold, "A Normal Form for Elliptic Curves," *American Mathematical Society,* v.44, 2007, pp.393-422.

[6] Gaj,Kris, Soonhak Kwon, Patrick Baier, and 4 others, "Implementing the Elliptic Curve Method of Factoring in Reconfigurable Hardware," citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.424&rep

[7] Hoffstein, Jeffrey, Jill Pipher and Joseph Silverman, *An Introduction to Mathematical Cryptography, Springer,* New York, 2008.

[8] Meulenaer,Giacomo de, Francois Gosset, Guerric Meurice de Dormale, and Jean-Jacques Quisquater, "Elliptic Curve Factorization Method: Towards Better Exploitation of Reconfigurable Hardware," SHARCS Workshop Record, 2007.

[9] Montgomery, Peter, "An FFT Extension of the Elliptic Curve Method of Factorization," UCLA dissertation, 1992.

[10] Montgomery, Peter, "Speeding up the Pollard and Elliptic Curve Methods of Factorization," *Math Comp.* v.48, 1987, pp. 243-264.

[11] Okeya, Katsuyuki, Hiroyuki Kurumatani and Kouichi Sakurai, "Elliptic Curves with the Montgomery-Form and Their Cryptographic Applications,"

[12] Pelzl,Jan, Martin Simka. Thorsten Kleinjung, Milos Drutarovsky, Victor Fischer and Christof Paar, "Area-Time efficient Hardware Architecture for Factoring Integers with the elliptic Curve Method," *Information Security, IEE Proceeding* v.152, 2005, pp.67-78.

[13] Pomerance, Carl and Richard Crandall, *Prime Numbers a Computational Perspective, 2nd ed. Springer*, New York, 2005.

[14] Washington, Lawrence, *Elliptic Curves Number Theory and Cryptography,* CRC Press, Boca Raton, 2003.

[15] Zimmermann, Paul and Bruce Dodson, "20 Years of ECM," *Springer-Verlag Berlin Heidelberg,* 2006, pp.525-542.

APPENDIX A

CODE

The following function `bisrho` approximates $u(e)$ discussed in section (5.2), compliments of Dr. Paul Duvall.

```
bits:=func<n|Ceiling(Log(2,n))>;
smooth:=func<n|bits(ff[#ff][1]) where ff:=Factorization(n)>;
bisrho:=function(tar)
x:=1.0;y:=1.0;
while DickmanRho(y) ge tar do y:=y+0.5; end while;

r:=DickmanRho(y);
dist:=Abs(tar-r);
z:=y;
while dist gt 0.000001 do z:=x+(y-x)/2;
r:=DickmanRho(z);
if r lt tar then y:=z;else x:=z;end if;
dist:=Abs(tar-r);
end while;

return z;      //where DickmanRho(z) approximately tar.
end function;
```

The next code was used in testing the speed of Montgomery addition and the two different parameters of the Edwards curves. It was run using Magma [4]. Note that the same addition chain is being used in each multiplication algorithm.

```
//EDWARDS VS MONTGOMERY
```

```
//This function takes an Edwards curve and makes a equivalent Montgomery curve.
E2M:=function(xe,ye,n) //(xe,ye) point on eddie, n field order
k:=GF(n);
xe:=k!xe;  ye:=k!ye;
pe:=[xe,ye,1];
d:=(xe^2+ye^2 -1)/(xe^2*ye^2);  //define EE: (y^2 +x^2)z^2 = z^4+ dx^2y^2

//Makes equiv Montgomery curve EM: BY^2Z=X^3 + AX^2Z + XZ^2
B:= 1/(1-d);  A:=(2*(1+d))/(1-d);
xm:=(1+ye)/(1-ye); ym:=(2*(1+ye))/(xe*(1-ye));
pm:=[xm,1]; // equiv pt on Montgomery
return <pe,d,pm,A,B>;
//<point on Edwards, d of Edwards, point on Montgomery, A,B defining Montgomery>
end function;

//Montgomery Addition Formula
madd:=function(P,Q,G,A)
// Adds P+Q, where G = P-Q, A is the parameter that defines the curve
Amul:=(A+2)/4;
  xn,zn := Explode(P);
  xm,zm := Explode(Q);
  x0,z0 := Explode(G);
if P eq Q then
  x1:=(xn+zn)^2*(xn-zn)^2;
  z1:=(4*xn*zn)*((xn-zn)^2 + (Amul*4*xn*zn));
else
```

```
  x1:=z0*((xm-zm)*(xn+zn)+(xm+zm)*(xn-zn))^2;
  z1:=x0*((xm-zm)*(xn+zn)-(xm+zm)*(xn-zn))^2;
end if;
return [x1,z1];
end function;

//Montgomery Multiplication Formula
mmul:=function(G,n,A)
//(Point, multiple you want of point, A the parameter of curve)
 k:=Intseq(n,2);t:=#k;
 R0:=G;R1:=madd(G,G,G,A);
 j:=t-1;
 while j ne 0 do
  if k[j] eq 0 then R1:=madd(R0,R1,G,A); R0:=madd(R0,R0,G,A);
   else
   R0:=madd(R0,R1,G,A); R1:=madd(R1,R1,G,A);
  end if;
 j:=j-1;
end while;
 return R0;
end function;

//Edwards formulas
eadd:=function(p1,p2,d)     //Adds p1+p2, where d defines the curve
X1,Y1,Z1 := Explode(p1);
X2,Y2,Z2 := Explode(p2);
```

```
A:=Z1*Z2; B:=A^2; C:=X1*X2; D:=Y1*Y2;
E:=d*C*D; F:=B-E; G:=B+E;

X3:=A*F*((X1+Y1)*(X2+Y2)-C-D);
Y3:=A*G*(D-C);
Z3:=F*G;

p3:=[X3,Y3,Z3];
 return p3;
end function;

emul:=function(G,n,d)
 //Input (Point, multiple of point you want, d parameter of the curve)
 k:=Intseq(n,2);t:=#k;
 R0:=G;R1:=eadd(G,G,d);
 j:=t-1;
 while j ne 0 do
  if k[j] eq 0 then R1:=eadd(R0,R1,d); R0:=eadd(R0,R0,d);
   else
   R0:=eadd(R0,R1,d); R1:=eadd(R1,R1,d);
  end if;
 j:=j-1;
end while;
 return R0;
end function;
```

```
//Edwards formulas2 without defining separate variables.
e2add:=function(p1,p2,d) //Adds p1+p2, where d defines the curve
Xp,Yp,Zp := Explode(p1);
Xq,Yq,Zq := Explode(p2);

Xr:=(Zp^3*Zq^3-d*Xp*Xq*Yp*Yq*Zp*Zq)*((Xp+Yp)*(Xq+Yq)-(Xp*Xq+Yp*Yq));
Yr := (Zp^3*Zq^3 + d*Xp*Xq*Yp*Yq*Zp*Zq)*(Yp*Yq-Xp*Xq);
Zr := (Zp^4*Zq^4 - d^2*Xp^2*Xq^2*Yp^2*Yq^2);

p3:=[Xr,Yr,Zr];
return p3;
end function;

e2mul:=function(G,n,d)
//Input (Point, multiple of point you want, d parameter of the curve)
 k:=Intseq(n,2);t:=#k;
 R0:=G;R1:=e2add(G,G,d);
 j:=t-1;
 while j ne 0 do
  if k[j] eq 0 then R1:=e2add(R0,R1,d); R0:=e2add(R0,R0,d);
   else
   R0:=e2add(R0,R1,d); R1:=e2add(R1,R1,d);
  end if;
 j:=j-1;
end while;
```

```
 return R0;
end function;

//This function calculates the times of emul2, emul and mmul
//Returns the total time in that order.
race:=function(l) //l #number of loops to take
mtime:=0; etime:=0; e2time:=0;

for i in [1..l] do
 //choses random parameters for Edwards
n:=RandomPrime(100);  e:=Random(1,10);
k:=GF(n^e);
ex:=Random(1,(n^e -1));
ey:=Random(1,(n^e -1));

//gets equiv Montgomery curve
j:= E2M(ex,ey,n^e);
pe:=j[1]; d:=j[2]; pm:=j[3]; A:=j[4]; B:=j[5];

m:=Random(10000,50000);
//Edwards time
t:=Cputime();
mpe:=emul(pe,m,d);
etime:=etime+Cputime(t);
//Edwards time 2
t:=Cputime();
```

```
mpe2:=e2mul(pe,m,d);
e2time:=e2time+Cputime(t);
//monty time
t:=Cputime();
mpm:=mmul(pm,m, A);
mtime:=mtime+Cputime(t);
end for;
 return <e2time,etime,mtime>;
end function;
```